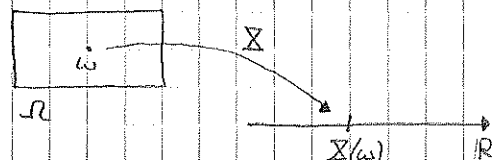


Def: En stokastisk variabel (även kallad s.v. eller slumpvariabel) X är en funktion från Ω till den reella tallingen $\mathbb{R}$.



En stokastisk variabel kopplar alltså ett tal $X(\omega)$ till varje utfall ω av ett slutförsök. Talet $X(\omega)$ skrivs oftast kortfattat som X och detta är anledningen till att X benämns stokastisk variabel och inte stokastisk funktion. Stokastiska variabler kan vara diskreta eller kontinuerliga. Födelsemånaden för en godtyckligt vald person är en diskret s.v. (med värden $1, 2, \dots, 12$) och längden hos en glödlampa är en kontinuerlig s.v. (med värden på intervallet $(0, \infty)$).

Def: En s.v. X sägs vara diskret om den antar ändligt eller uppräkneligt oändligt antal värden. Värdemängden kan skrivas $\{k_1, k_2, \dots\}$. För en diskret s.v. definieras sannolikhetsfunktionen p_X av $p_X(k) = P(X = k)$ för $k = k_1, k_2, \dots$

Sannolikhetsfunktionen beskriver hur sannolika de olika värdena på X är och kan åskådliggöras med stolpsdiagram. Oftast antar diskreta s.v. heltalsvärden och sannolikhetsfunktionen har då följande egenskaper:

$$0 \leq p_X(k) \leq 1 \quad \text{för alla } k$$

$$\sum_{k=-\infty}^{\infty} p_X(k) = 1$$

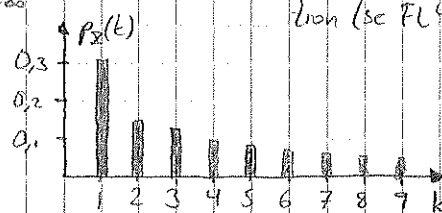
$$P(a \leq X \leq b) = \sum_{k=a}^b p_X(k)$$

$$P(X \leq a) = \sum_{k=-\infty}^a p_X(k) := F_X(a)$$

Denna funktion kallas för delningsfunktion (se FL 4)

Ex: Benfords lag beskriver hur förstasiffrorna i många numeriska datamängder (ex. vis städens invånarantal, tal i årsredovisningar, fysikaliska konstanter) är fördelade. Lagen kan användas för att testa om siffrorna i en datamängd kan ha blivit manipulerade. Sannolikhetsfunktionen är

$$p_X(k) = \log_{10} \left(\frac{k+1}{k} \right) = \log_{10}(k+1) - \log_{10}(k) \quad \text{för } k = 1, 2, \dots, 9.$$



Notera att $p_X(k) \geq 0$, ty $\frac{k+1}{k} \geq 1$ och att $\sum_{k=1}^9 p_X(k) = \sum_{k=1}^9 (\log_{10}(k+1) - \log_{10}(k))$

En kontinuerlig s.v. kan anta alla värden i ett eller flera intervall på $\mathbb{R}$. Utfallen ligger oändligt tätt, så inget enskilt utfall har positiv sannolikhet. Vi kan därför inte definiera sannolikhetsfunktionen som för diskreta s.v.

Def: En s.v. X sägs vara kontinuerlig om det existerar en icke-negativ funktion f_X så

$$P(X \in A) = \int_A f_X(x) dx \quad \text{för alla mängder } A \subset \mathbb{R}$$

Funktionen f_X kallas täthetsfunktionen för X . Notera att $\int_{-\infty}^{\infty} f_X(x) dx = 1$.

Sannolikheten för en händelse ges för en kontinuerlig s.v. av integration av täthetsfunktionen istället för summation av sannolikhetsfunktionen. Täthetsfunktionen kan inte direkt tolkas som en sannolikhet, men det gäller att

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx, \quad (*)$$



så sannolikheten är lika med arean under kurvan. Det spelar ingen roll om vi använder $<$ eller $\leq$ i (*), eftersom $P(X=c) = 0$ för alla $c \in \mathbb{R}$ om X är en kontinuerlig s.v.

Vi ska nu undersöka några diskreta sannolikhetsfördelningar, dvs. ofta förekommande klasser av sannolikhetsfunktioner.

Def: En s.v. X sägs vara likformigt fördelad på $\{1, 2, \dots, m\}$ om $p_X(k) = \frac{1}{m}$ för alla $k \in \{1, 2, \dots, m\}$. (Antalet prickar vi ett tärningskast är ett exempel på en likformigt fördelad s.v. på $\{1, 2, \dots, 6\}$.)

Def: En s.v. X sägs vara tvärpunktsfördelad om $p_X(a) = p$ och $p_X(b) = 1 - p$, där $p \in (0, 1)$ och $a, b \in \mathbb{R}$. Om $a = 1$ och $b = 0$ sägs X vara Bernoulli-fördelad. Detta betecknas $X \in \text{Be}(p)$. (Uttallet vid ett experiment som lyckas ($X=1$) med sannolikhet p och annars misslyckas ($X=0$) kan beskrivas med en Bernoullifördelad s.v.)

Def: En s.v. X sägs vara för-första-gången-fördelad om $p_X(k) = (1-p)^{k-1} p$, där $p \in (0, 1)$ och $k = 1, 2, \dots$. Detta betecknas $X \in \text{ffg}(p)$. (Om ett experiment som lyckas med sannolikhet p upprepas så är den gång i ordningen då det lyckas första gången ffg-fördelad.)

Vi härleder den viktiga binomialfördelningen genom att ytterligare undersöka exemplet med ett försök som lyckas med sannolikhet p . Vilken är sannolikheten att om försöket upprepas n gånger så kommer det att lyckas k gånger? Sannolikheten att försöket först lyckas k gånger och sedan misslyckas $n-k$ gånger är $p^k (1-p)^{n-k}$. Detta är sannolikheten för varje sekvens av n försök som innehåller exakt k lyckade försök. Totalt finns $\binom{n}{k}$ sådana sekvenser (välj ut de k omgångarna då försöket lyckas). Den totala sannolikheten för k lyckade försök är således $\binom{n}{k} p^k (1-p)^{n-k}$.

Def: En s.v. X sägs vara binomialfördelad om $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$, där $p \in (0, 1)$ och $k = 0, 1, \dots, n$. Detta betecknas $X \in \text{Bin}(n, p)$.

Binomialsatsen säger att $(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$, så med $x=p$ och $y=1-p$ fås

$$\sum_{k=0}^n p_X(k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1 \quad \text{som önskat.}$$

BILD 1 $\rightarrow$ Bilden visar sannolikhetsfunktionen $p_X(k)$ för fyra olika binomialfördelningar med $n=20$, nämligen $p=0,1$, $p=0,25$, $p=0,5$ och $p=0,75$. Som synes sker en förskjutning mot större värden när p ökas.

Ex: Låt X vara antalet sexor som fås vid 20 tärningskast. Då är $X \in \text{Bin}(20, \frac{1}{6})$. Sannolikheten att få minst 8 sexor vid 20 tärningskast är

$$P(X \geq 8) = \sum_{k=8}^{20} \binom{20}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{20-k} \approx 1,1\%$$

Vi undersöker nu fallet då sannolikheten att experimentet lyckas är mycket liten. Låt alltså p vara litet, men låt samtidigt n vara stort, så att $\mu = np$ har storleksordning ett. Då gäller:

$$\begin{aligned} p_X(k) &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k} \\ &= \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^n \underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\approx 1 \text{ om } k \ll n} \underbrace{\left(1 - \frac{\mu}{n}\right)^{-k}}_{\approx 1 \text{ om } k \ll n} \approx \frac{\mu^k}{k!} e^{-\mu} \end{aligned}$$

Def: En s.v. X sägs vara Poissonfördelad om $p_X(k) = \frac{\mu^k}{k!} e^{-\mu}$ för $\mu > 0$ och $k = 0, 1, \dots$. Detta betecknas $X \in \text{Po}(\mu)$.

Poissonfördelning används då ett försök med litet sannolikhet att lyckas upprepas många gånger. För $n > 10$ och $p < \frac{1}{10}$ är approximationen $\text{Bin}(n, p) \approx \text{Po}(np)$ god.